E7.6-10268.

LARS Information Note 012176

NASA CR-
147518

RESULTS FROM THE CROP IDENTIFICATION TECHNOLOGY ASSESSMENT FOR REMOTE SENSING (CITARS) PROJECT

R. BIZZELL, F. HALL, A. FEIVESON,
M. BAUER, B. DAVIS, W. MALILA, AND
D. RICE

(E76-10268) RESULTS FROM THE CROP IDENTIFICATION TECHNOLOGY ASSESSMENT FOR REMOTE SENSING (CITARS) PROJECT (Purdue Univ.) 40 p HC $4.00 CSCL 02B

N76-21635

Unclas
G3/43 00268

The Laboratory for Applications of Remote Sensing

Purdue University, West Lafayette, Indiana

1976

LARS Information Note 012176

Results from the Crop Identification Technology Assessment for Remote Sensing (CITARS) Project

by

R. M. Bizzell, F. G. Hall and A. H. Feiveson
National Aeronautics and Space Administration
Earth Observation Division (EOD)
Johnson Space Center
Houston, Texas

M. E. Bauer and B. J. Davis
Laboratory for the Applications of Remote Sensing (LARS)
Purdue University
West Lafayette, Indiana

W. A. Malila and D. P. Rice
Environmental Research Institute of Michigan (ERIM)
Ann Arbor, Michigan

Laboratory for Applications of Remote Sensing
Purdue University
West Lafayette, Indiana

The LARS/Purdue University portion of the CITARS project was supported by the National Aeronautics and Space Administration under contract NAS9-14016.

This Information Note presents two papers which were presented at the Ninth and Tenth International Symposia on Remote Sensing of Environment, Ann Arbor, Michigan, April 1974 and October 1975. The papers summarize the objectives, experimental procedures, and results of the CITARS project.

TABLE OF CONTENTS

RESULTS FROM THE CROP IDENTIFICATION TECHNOLOGY ASSESSMENT FOR REMOTE SENSING (CITARS) PROJECT*

R. M. Bizzell, F. G. Hall and A. H. Feiveson
National Aeronautics and Space Administration
Earth Observations Division (EOD)
Johnson Space Center
Houston, Texas

M. E. Bauer and B. J. Davis
Laboratory for the Applications of Remote Sensing (LARS)
Purdue University
West Lafayette, Indiana

W. A. Malila and D. P. Rice
Environmental Research Institute of Michigan (ERIM)
Ann Arbor, Michigan

ABSTRACT

The CITARS task design and objectives are reviewed and final results are presented, together with conclusions and recommendations. It was found that several factors had a significant effect on crop identification performance: (a) crop maturity and site characteristics, (b) which of several different single-date automatic data processing procedures was used for local recognition, (c) nonlocal recognition, both with and without preprocessing for the extension of recognition signatures, and (d) use of multidate (multitemporal) data. It also was found that classification accuracy for field center pixels was not a reliable indicator of proportion estimation performance for whole areas, that bias was present in proportion estimates, and that training data and procedures strongly influenced crop identification performance.

1. INTRODUCTION AND OBJECTIVES

CITARS (Crop Identification Technology Assessment for Remote Sensing) was a joint task to quantify the crop identification performance (CIP) achievable with several automatic data processing (ADP) techniques operating on remote sensor data. It was conducted from April 1973 to April 1975. Participants were the Earth Observations Division (EOD) of the Johnson Space Center, the Environmental Research Institute of Michigan (ERIM), the Laboratory for Applications of Remote Sensing (LARS) of Purdue University, and the Agricultural Stabilization and Conservation Service (ASCS) of the U.S. Department of Agriculture. The CITARS task design was presented in Ref. 1 which also includes objectives, analysis methodology, experimental procedures, description of ADP procedures, and first results of CITARS. A more extensive description and documentation of the entire project can be found in Ref. 2. The remainder of this section presents a brief overview of the CITARS task design and objectives.

The principal assessments made were of crop identification performance for corn and soybeans in six sites in Illinois and Indiana. The remote sensor data analyzed were collected by the ERTS-1 (now called LANDSAT-1) multispectral scanner (MSS) periodically throughout the 1973 growing season. The ADP procedures compared were predefined at EOD, ERIM, and LARS in such a way as to minimize subjective analyst judgment and intervention in the crop identification process.

Specific objectives of CITARS included performance comparisons to answer the following questions:

---

* **Proceedings of the Tenth International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, October 6-10, 1975.**

1. How does corn and soybean identification performance vary with time during the growing season?

2. How does crop identification performance vary among different geographic locations having different soils, weather, management practices, crop distributions, and field sizes?

3. How much variation in crop identification performance is observed among the different data processing procedures?

4. Can signal statistics acquired from one time or location be used to identify crops at other locations and/or times?

5. Does use of radiometric preprocessing extend the use of training statistics and/or increase nonlocal crop identification performance?

6. Does use of multitemporal data increase crop identification performance?

To accomplish the objectives, five major tasks had to be completed. These were: (1) acquisition and preparation of an ERTS data set with ancillary data sufficient to support the experimental objectives and design, (2) computer-aided processing of this data set with the selected classification algorithms and procedures, (3) quantification of the crop identification performances in a manner which would permit a quantitative evaluation of the ability of these procedures to satisfy agricultural applications requirements, (4) a statistical analysis to quantitatively evaluate the impact of major factors known to affect crop identification performance, and (5) an interpretation of the results to ascertain the underlying physical factors responsible for the results, to draw inferences as to the status of the technology as it relates to agricultural applications, and to make recommendations as to where the technology must be strengthened.

2. KEY TECHNICAL ACCOMPLISHMENTS

As reported earlier [1] an ERTS data set with supporting ancillary data was acquired and prepared which met the requirements of the CITARS experimental design, except for completeness of ERTS coverage. Assembly of the data set included: (1) acquisition of crop identification and other agronomic ground truth data by ASCS, (2) acquisition and interpretation of color infrared aerial photography to extend the field identification data acquired by ASCS to additional fields, (3) registration and geometric correction of multitemporal ERTS MSS data for the test segments, and (4) location of field and section coordinates in the ERTS data. In addition, repeatable, analyst-independent ADP procedures had to be defined and documented and measures of crop identification performance had to be determined.

Each of the six 5x20-mile (8x32-km) sites was located in an overlap zone of ERTS so that coverage was available on two successive days on each 18-day ERTS cycle. Of 72 potential data sets from late June through late September, only 15 were sufficiently cloud free for use in the analysis (See Table I).

Periodic crop observations of fields used to train the classifiers were made by USDA/ASCS throughout the growing season. These fields were found in 20 quarter sections (0.8x0.8 km), each located randomly within a 1x5-mi (1.6x8-km) strip in the site. Photointerpretation of multidate aerial photography was successfully used to increase the size of the data base. The photointerpretation data in 20 full sections (1.6x1.6 km), randomly selected throughout the site, were used to evaluate ERTS data classification accuracy in field centers. In addition, crop area proportion measurements were made from the aerial photography and used to evaluate proportion estimates derived from pixel-by-pixel classification of ERTS data. The accuracy of photointerpretation was tested in 223 ASCS-visited fields which were not revealed to the photointerpreters. The results, summarized in Table II, were judged to be of sufficient accuracy to warrant use of the photo-interpreted field identifications for evaluation of the ERTS data processing results.

Multiple passes of ERTS data were registered with an average root mean square (rms) error of less than one-half pixel, enabling multitemporal classifications of the data and eliminating the need to locate field and section coordinates in each ERTS scene.

The need to maximize the number of pure pixels selected from the relatively small-sized fields present in several of the segments made selection of field coordinates more difficult than expected. Manual methods were found to be inadequate for the job and a computer-aided method of transforming digitized photomap coordinates to ERTS line and column coordinates was used [1]. The use of the latter method is recommended for future projects requiring precise definition of ERTS data coordinates.

A key task prior to the start of ERTS data classifications was to define and document data analysis procedures which were repeatable, easily followed, and yet incorporated the judgment and skill of experienced analysts. Although there was concern that crop identification performance might be reduced by restricting analyst decisions, it was necessary that variability due to analysts be minimized if meaningful comparisons of results were to be made. Limited tests with the LARS ADP procedures indicated that, for the CITARS data set, the procedures used produced results comparable to those obtainable using procedures with considerably more analyst interaction.

To evaluate crop identification performance, three categories or classes of data were defined. The first two were the major crops, corn and soybeans, while the third, called "other", included all other ground covers. Analysis of wheat recognition early in the year was attempted, but the amount of wheat in the segments was too small and the reliability of its photointerpretation too low to support meaningful conclusions.

An important accomplishment of CITARS was the use of quantitative measures of crop identification performance and statistical evaluation of results. The statistical analyses consisted of analyses of variance and blocked rank tests for comparisons involving factors such as ADP procedures, location, acquisition date, and use of preprocessing. Two variables, average conditional classification accuracy of "pure" field center pixels and the rms error of proportion estimates for entire sections, were used as measures of classification performance. Section-to-section variability was used in analyses of variance to determine if differences among ADP procedures, segments, times, etc. were significant. The analyses of variance revealed several significant differences, but the power of many of the significance tests was limited due to missing data, the amount of variability present, and the failure of the dependent variable to adequately describe performance for a section independently of the composition of that section, despite the use of a normalizing transformation. Continued use and development of these tools for remote sensing experiments are recommended.

3. RESULTS AND DISCUSSION

The statistical analyses provided a key to the quantitative assessment of remote sensing technology for crop identification, for both field center pixels and crop area estimation. Previous results were confirmed in some instances, while in others unanticipated results led to reconsiderations and new insights into certain aspects of the technology. This section summarizes the major results from the CITARS experiments, as related to the six specific objectives which were expressed as questions regarding the effects of various factors on crop identification performance.

3.1 EFFECTS OF TIME DURING THE GROWING SEASON

The time of ERTS data acquisition during the growing season was found to be an important factor influencing crop identification performance, because of the phenological development cycles or crop calendars of the major ground covers. The peak accuracy for field center classification was 75-80% correct in mid-August, as shown in Fig. 1. At this time, the variability within the major crops (corn and soybeans) was low and the amount of ground cover was high.

The solid line in Fig. 1 represents the expected performance for the average of all single-date procedures, assuming no interaction between the factors: site and time. The use of a non-interactive model for computing expected performance as a function of time was necessitated by the fact that only one site had ERTS data for more than two of the six time periods. The individual points marked on the graph represent actual performances by the various procedures. The variability present is an indication that factors other than time also influenced field center classification performance.

A similar expected performance time profile also was calculated for proportion estimates over the aggregation of whole test sections. This profile showed roughly the same rms proportion error for all time periods, except mid-July when the error was substantially greater. In mid-July, the variability among corn fields and among soybean fields was high, and the amount of ground cover was low. We do note, however, that variability in performance among procedures at any given time was much greater for proportion estimation than in the case of field center classification.

3.2 EFFECTS OF SITE

Missing data again hampered the analysis when comparisons were made between sites. Nevertheless, proportion estimation accuracy was found to be much more site dependent than was field center classification accuracy, when expected responses were computed. The only major site characteristics which were found to be correlated with proportion estimation accuracy were average

field size and proportion of corn and soybeans in the segment. As shown in Fig. 2, proportion estimation errors were smallest for the site with the largest average field size. Similarly, the proportion estimation error was found to be smallest for the site with the greatest percentage of corn and soybeans among its ground cover types.

The correlation of field size and the accuracy of crop proportion estimates is attributed primarily to the decrease in the percent of pixels containing mixtures of crops as field size increases. In addition, it has been observed that large fields tend to be more uniform, and areas having larger fields have relatively fewer fields of "other" covers. The influence of these factors on crop identification performance will be discussed further in Sec. 3.3.

3.3 EFFECTS OF ADP PROCEDURE ON LOCAL RECOGNITION WITH SINGLE-TIME DATA

EOD, ERIM, and LARS each defined a principal ADP procedure which was used for the major comparisons of crop identification performance using data from single ERTS passes; alternate procedures also were defined and tested by ERIM and LARS. In this section these procedures are compared for local recognition, that is, when the test data were located in the same site and ERTS pass as the training data. Nonlocal recognition is considered in Sec. 3.4. Identical training field, test field, and test section coordinates were used with all procedures.

3.3.1 PRINCIPAL PROCEDURES. Major differences were found in results for the three principal ADP procedures. Local performance results with these procedures are summarized in Table III, where overall average performance figures are given, as well as the number of specific analysis of variance (ANOVA) comparisons for which significant differences were detected (out of a total of ten comparisons). The ERIM/SP1 procedure was best for field center classification, while LARS/SP1 was the most consistent for whole area proportion estimation and had the lowest average rms proportion error. These results indicate that field center classification of pure pixels which has commonly been used to evaluate crop identification performance is not a reliable indicator of the accuracy of proportion estimates for whole areas.

Primary differences between the ADP procedures lie in the training procedures and decision rules used. Yet, there are some characteristics which they share that can contribute to the observed results.

First, there is inherent bias in proportion estimates based on aggregated counts of maximum likelihood pixel-by-pixel classifications. Bias exists because the expected performance of a classifier depends on the true crop proportions present, as well as on its performance matrix for individual pixels. As can be seen in Table III, all three principal procedures consistently underestimated the proportion of "other" in the test data. Furthermore, the expected rms error in proportion estimation was found to be correlated with the percent of other in a test site.

Second, the whole areas included boundary pixels which contain mixtures of two or more ground covers. Mixture pixels were determined to be a major source of biased proportion estimates by a special analysis, as well as by the fact that expected proportion errors tended to be largest in segments with the smallest average field size (Fig. 2) and, therefore, with the greatest number of mixture pixels.

The three principal procedures tested differed in two ways. Both LARS/SP1 and EOD/SP1 used a clustering procedure to define training statistics (usually several classes for each major crop) and employed a quadratic decision rule. ERIM/SP1, on the other hand, formed a single signature for each major crop, used a variable number of signatures for "other", and used a linear decision rule. The differences in performance among the three procedures were determined to be due to the method of training rather than the decision rule used since similar results (high ranking for field center recognition and low ranking for whole area proportion estimates) were obtained for ERIM/SP2, a quadratic decision rule classifier, which used the same signatures as ERIM/SP1. It was observed that the disparity between rankings for the two types of performance for the ERIM procedure tended to be reduced or eliminated when within-corn and within-soybean variations were smallest and the procedure selected greater numbers of other-class signatures.

An attempt was made to correct for classifier bias in proportion estimates by using the performance matrix for field-center pixels, but the attempt was unsuccessful both on a section-by-section and an aggregated segment basis.

3.3.2 ALTERNATE PROCEDURES. A linear decision rule optimized on a class-pairwise basis was used in ERIM's principal procedure and a quadratic decision rule, similar to those of LARS/SP1 and EOD/SP1, in its alternate procedure. It was found that the accuracy of results with the linear rule were equal to or better than those with the quadratic rule using the same signatures. Resource constraints of CITARS did not permit similar comparisons with signature sets obtained by a different procedure, but such or similar comparisons are recommended.

Another CITARS result which, on the surface, seems surprising is the lack of improvement of LARS/SP2 (non-equal major class prior probabilities) over LARS/SP1 (equal prior probabilities). Theoretically, apart from boundary pixels, the Bayesian classifier should produce its minimum error rate when "correct" values for the frequency of occurrence of each spectral class are utilized as parameters in the classification rule. LARS/SP2 included a procedure for estimating the prior probability of each spectral class based on existing agricultural statistics. For CITARS classifications the LARS/SP2 procedure utilizing unequal prior probabilities did not produce an improvement over assuming them to be equal. This is attributed in part to the fact that the agricultural statistics used were at a county level only, differing by as much as 20 percent from the true proportions in the test sections which were subsets of the county and not randomly located within it. Boundary pixels are another possible cause. We do not believe that use of prior probability information in the form of class weights should be discouraged, based solely on the CITARS analysis since it does not constitute a definitive test. Instead, it is recommended that further tests be made to determine the sensitivity of maximum likelihood classifier to class weights.

In other experiments LARS showed that significant differences in classification performance can be obtained with different training sets and that training set size alone does not determine the adequacy of a training set. These results and results discussed earlier point out the dependence of crop identification performance on the development of training statistics.

3.4 EFFECTS OF NONLOCAL RECOGNITION AND PREPROCESSING

In nonlocal recognition, the training statistics are used to recognize data from a different location and/or a different ERTS pass. Such procedures are desirable and/or necessary in order to reduce the cost of obtaining ground identification information for training classifiers in operational applications. The effect of nonlocal recognition on performance with the single-time procedures was evaluated for 20 pairings of the 15 data sets.

Comparisons of classification performance indicated that average field center performance for the three principal procedures in nonlocal recognition was reduced by 22 percent of that obtained locally. For whole area proportion estimates, the average rms error of nonlocal estimates was 23 percent greater than that for local estimates. The degradation associated with nonlocal classification performance was shown to be correlated ($r = -0.77$) with differences in atmospheric optical depth (a measure of haze level) between the training and recognition segments. Other differences present in the data sets were those of soil type, agricultural practice, crop maturity, scene composition, training data selection, and MSS scan angle, all of which can affect the representativeness of signatures. The results clearly indicate problems in successfully applying training statistics to different locations and/or ERTS passes.

One way of extending the realm of applicability of signatures is to transform them radiometrically so they better represent observation conditions at recognition segments. Preprocessing with a mean level adjustment algorithm (ERIM/PSP1), a relatively simple preprocessing algorithm, was found to be of some help in improving nonlocal recognition performance. Overall, the preprocessing procedure ranked above the three principal procedures for both whole area proportion estimation and field center classification; this ranking was statistically significant for field centers but not for whole areas. Preprocessing cut roughly in half the degradation in mean field center performance and substantially reduced the correlation between optical depth differences and field center performance (from $r = -0.77$ to $r = -0.28$), but was not consistent in its performance, especially for whole areas.

The mixed results obtained in specific analyses of variance indicate that differences in composition of training and test areas also are important factors affecting nonlocal recognition. Additional research is required to improve upon the signature adjustment algorithm tested and to better account for spectral variability due to scene composition. A limited test of a more complex signature extension algorithm at ERIM, in an effort supplementary to CITARS, indicated that improved results are possible.

3.5 EFFECTS OF MULTITEMPORAL DATA

One CITARS segment (Fayette) had several clear ERTS overpasses which were spatially registered and then analyzed and processed multitemporally with the EOD/MSP1 procedure. Significant increases in crop identification performance were obtained, compared to the best single-date performance. Use of multitemporal data increased field center classification accuracy from 81 percent to 89 percent correct and halved the rms error in proportion estimation. These substantial improvements in performance were obtained for this one segment by using basically the same data analysis procedures as for single-date data; nevertheless, new analysis procedures taking into account the increased complexity of multitemporal scenes will need to be researched and developed.

Although use of multitemporal data requires a more complex data processing system (registration, increased data base size, and more complex data analysis procedures), the increased complexity may well be justified by increased performance.

3.6 RELATION OF CROP AND SENSOR CHARACTERISTICS

Two key factors influencing crop identification with remote sensor data are (1) the nature of the spectral variation among and within the classes to be identified and (2) the capability of the sensor to measure the spectral variation. An understanding of the relationship of these factors may help explain the levels of crop identification performance obtained in CITARS. In several instances it was found that accurate identification of corn, soybeans, and "other" was not possible even when all the fields analyzed were used to train the classifier. This may have been due to a lack of differences in the spectral characteristics of the three classes or to an inability of the ERTS MSS to resolve and precisely measure the differences present. The latter is suspected to account for at least a part of the problem since crop classifications made during the 1971 Corn Blight Watch Experiment [3] using MSS data with more spectral bands, narrower bands, and greater sensitivity and dynamic range showed that these same cover types could be more accurately identified. Additional comparisons of ERTS and aircraft-acquired MSS or other high spectral resolution data such as will be available from the current LACIE (Large Area Crop Inventory Experiment) field measurements project [4] will be needed to verify this point.

4. CONCLUSIONS AND RECOMMENDATIONS

4.1 CONCLUSIONS

CITARS has provided a quantitative assessment of 1973-era technology for remote identification of major agricultural crops. The use of quantitative measures of classification performance and statistical evaluations of the results have been important parts of the technology assessment. The major conclusions from the CITARS experiments are:

1. Crop identification performance for corn and soybeans varied throughout the growing season, with field center classification accuracy being maximum in late August.

2. The probability of correct classification of field center pixels was not well correlated with and thus was not a reliable indicator of proportion estimation performance.

3. Proportion estimation accuracy was strongly correlated with both average field size and proportion of major crops in the segment, but field center classification accuracy was not. Boundary pixels containing two or more cover types were determined to be major contributors to the bias in proportion estimates.

4. The manner in which ground cover classes were selected and used to train the classifier strongly influenced the amount of bias in proportion estimates.

5. Probability of correct classification and proportion estimation accuracy both were decreased when training statistics developed for a different location or date were used.

6. A mean level adjustment algorithm for first order adjustments to training statistics used for nonlocal classifications increased the probability of correct classification of field center pixels, but did not improve proportion estimates for whole areas.

7. The use of multidate data improved both proportion estimation accuracy and probability of correct classification.

In addition it has been shown that relatively automatic data analysis procedures can be defined which produce repeatable results, are suited for processing relatively large data volumes, and incorporate, to a large degree, the judgment and expertise of experienced analysts.

4.2 RECOMMENDATIONS

CITARS provides valuable direction for future research and development of remote sensing technology and guidelines for the design of operational crop production survey systems utilizing remote sensing technology. Recommendations from CITARS include:

1. Continued use and development of quantitative measures of crop identification performance and statistical evaluation of classification results.

2. Continued development of improved methods for training classifiers.

3. Research and development of methods to improve the accuracy of proportion estimates for whole areas.

4. Further tests to determine the sensitivity of maximum likelihood classifiers to the use of prior probability information and of the linear classifier to different signature sets.

5. Additional research, development, and testing of two complementary approaches to nonlocal recognition, (a) more sophisticated preprocessing algorithms and (b) stratification of areas based on their similarity with respect to agricultural factors.

6. Development of data analysis procedures which account for the increased complexity of multitemporal data and take advantage of its potentially greater information content.

7. Additional comparisons of ERTS and other multispectral data sources to determine the adequacy of ERTS MSS in terms of the number, width and placement of its spectral bands, signal/noise ratio, sensitivity, dynamic range, and spatial resolution.

ACKNOWLEDGEMENTS

The authors wish to acknowledge and thank those others at EOD, ERIM, and LARS who contributed to the success of the CITARS effort through planning, implementation, execution, and/or analysis of the CITARS data and results. We also appreciate the cooperation of USDA/ASCS personnel in providing the field observations.

REFERENCES

1. Hall, F. G., Bauer, M. E., and Malila, W. A., First Results from the Crop Identification Technology Assessment for Remote Sensing (CITARS). Proc. Ninth Int. Symp. on Remote Sensing of Environment, Ann Arbor, Michigan, 1974, pp. 1171-1191.

2. Crop Identification Technology Assessment for Remote Sensing (CITARS), NASA Johnson Space Center, Houston, Texas, 1975.

   Vol. I Task Design Plan, by F. G. Hall. JSC-09384, Feb. 1975.

   Vol. II Ground-Truth Field Observations of the Agricultural Stabilization and Conservation Service, Parts 1-6, by J. W. Dietrich. JSC-09385, Feb. 1975.

   Vol. III The First Earth Resources Technology Satellite (ERTS-1) Multispectral Scanner Data Acquisition (and Supplement), by R. M. Bizzell. JSC-09386, Feb. 1975.

   Vol. IV Image Analysis. JSC-09387:
   Part 1 Crop Identification Extension, by J. W. Dietrich, Apr. 1975.
   Part 2 Evaluation (to be published)

   Vol. V The First Earth Resources Technology Satellite (ERTS-1) Data Preparation, by W. R. Simmons, D. M. Freeman, and M. E. Bauer. JSC-09388, March 1975.

   Vol. VI Data Processing at the Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, Indiana, by M. E. Bauer, T. K. Cary, B. J. Davis, and P. H. Swain, JSC-09389, July 1975.

   Vol. VII Data Processing at the Environmental Research Institute of Michigan, Ann Arbor, by W. A. Malila, D. P. Rice, and R. C. Cicone. JSC-09390, June 1975.

Vol. VIII Data Processing at the Earth Observations Division, Lyndon B. Johnson Space Center. JSC-09391:

Part 1 Huntington County, Indiana, by F. Johnson, March 1975.
Part 2 Shelby County, Indiana, by G. L. Gutschewski, April 1975.
Part 3 White County, Indiana, by R. L. Patterson, March 1975.
Part 4 Livingston County, Illinois
Part 5 Fayette County, Illinois, by R. L. Patterson, May 1975.
Part 6 Lee County, Illinois, by G. L. Gutschewski, April 1975.
Part 7 Multitemporal Analysis of Fayette County, Illinois, by J. G. Garcia, May 1975.

Vol. IX Statistical Analysis of Results, by B. J. Davis and A. H. Feiveson. JSC-09392, Sept. 1975.

Vol. X Interpretation of Results, by R. M. Bizzell, A. H. Feiveson, F. G. Hall, M. E. Bauer, B. J. Davis, W. A. Malila, and D. P. Rice. JSC-09393 (to be published).

3. Corn Blight Watch Experiment Final Report, Vol. III, Experiment Results, NASA Johnson Space Center, Houston, Texas, May 1974.

4. The Large Area Crop Inventory Experiment (LACIE) -- An Assessment After One Year of Operation, by R. B. MacDonald, F. G. Hall, and R. B. Erb. Proc. of Tenth Int. Symp. on Remote Sensing of Environment, Ann Arbor, Michigan, October 1975

TABLE I. ERTS DATA SETS ACCEPTED FOR CITARS PROCESSING

	ERTS OVERPASS CYCLE					
SITE	JUNE 26-30	JULY 14-18	AUG 1-5	AUG 19-23	SEPT 6-10	SEPT 24-28
1. HUNTINGTON, IND.		B				A
2. SHELBY, IND.					B	A
3. WHITE, IND.				B	A	
4. LIVINGSTON, ILL.		A	A			
5. FAYETTE, ILL.	B	A&B		A		
6. LEE, ILL.		B	B			

Key: A = First Pass of ERTS Coverage of Site
B = Second Pass of ERTS Coverage

TABLE II. COMPARISON OF CROP TYPE IDENTIFICATIONS MADE BY ASCS AND BY PHOTOINTERPRETATION

COVER TYPE	STATISTIC	ASCS TOTAL	PHOTOINTERPRETED TOTAL	PHOTOINTERPRETED CORRECT	PHOTOINTERPRETED COMMISSION ERROR
CORN	Number of Fields	50	51	46	5
	Percentage	100.0	102.0	92.0	10.0
	Number of Acres	1,181	1,197	1,165	32
	Percentage	100.0	101.3	98.6	2.7
SOYBEANS	Number of Fields	65	66	61	5
	Percentage	100.0	101.5	93.8	7.7
	Number of Acres	1,550	1,540	1,523	16
	Percentage	100.0	99.3	98.3	1.1
OTHER	Number of Fields	108	106	99	7
	Percentage	100.0	98.1	91.7	6.5
	Number of Acres	879	874	838	36
	Percentage	100.0	99.4	95.3	4.1

TABLE III. COMPARISON OF PRINCIPAL PROCEDURES FOR LOCAL RECOGNITION

A. MEAN CLASSIFICATION ACCURACY FOR FIELD CENTER PIXELS (15 CASES)

CLASS	LARS/SP1	ERIM/SP1	EOD/SP1
CORN	0.66	0.70	0.62
SOYBEANS	0.59	0.68	0.61
OTHER	0.5[illegible]	0.53	0.46
OVERALL	0.58	0.64	0.57

- FOR SEVEN ANOVA COMPARISONS WHERE SIGNIFICANT DIFFERENCES WERE DETECTED FOR FIELD CENTER PERFORMANCE, ERIM/SP1 RANKED FIRST IN SIX.

B. MEAN PROPORTION ESTIMATION BIASES AND RMS ERROR FOR WHOLE AREAS (15 CASES)

	CLASS	LARS/SP1	ERIM/SP1	EOD/SP1
BIAS FOR:	CORN	0.063	0.064	0.025
	SOYBEANS	0.033	0.059	0.081
	OTHER	−0.096	−0.124	−0.106
	RMS ERROR	0.095	0.150	0.108

- FOR EIGHT ANOVA COMPARISONS WHERE SIGNIFICANT DIFFERENCES WERE DETECTED FOR PROPORTION ESTIMATION, LARS/SP1 RANKED FIRST IN ONE AND SECOND IN SEVEN COMPARISONS, ERIM/SP1 FIRST IN THREE CASES, AND EOD/SP1 FIRST IN FOUR COMPARISONS.

FIGURE 1. VARIABILITY OF CROP IDENTIFICATION PERFORMANCE DURING GROWING SEASON

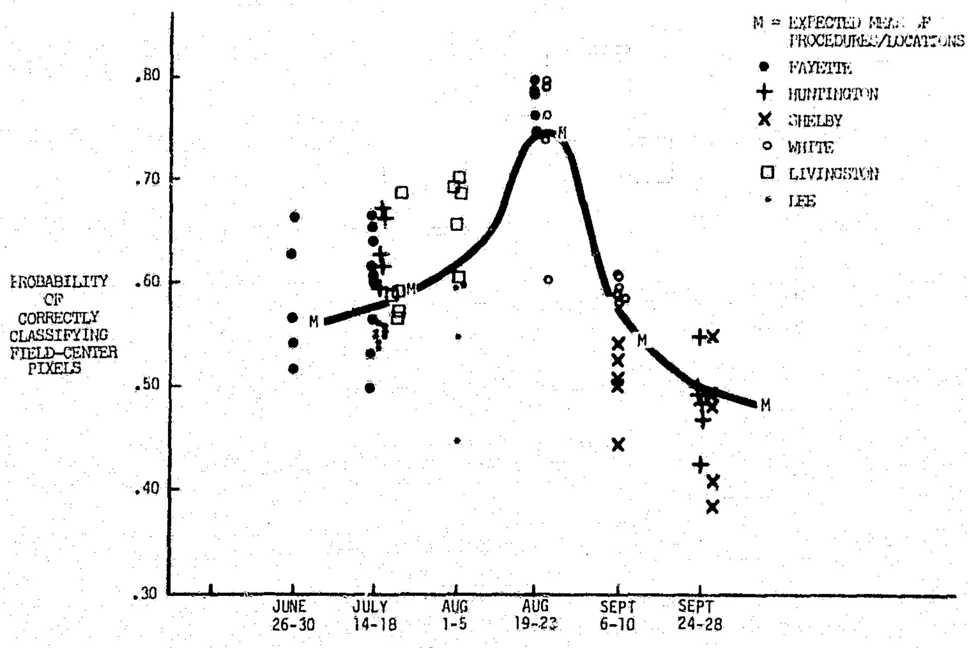
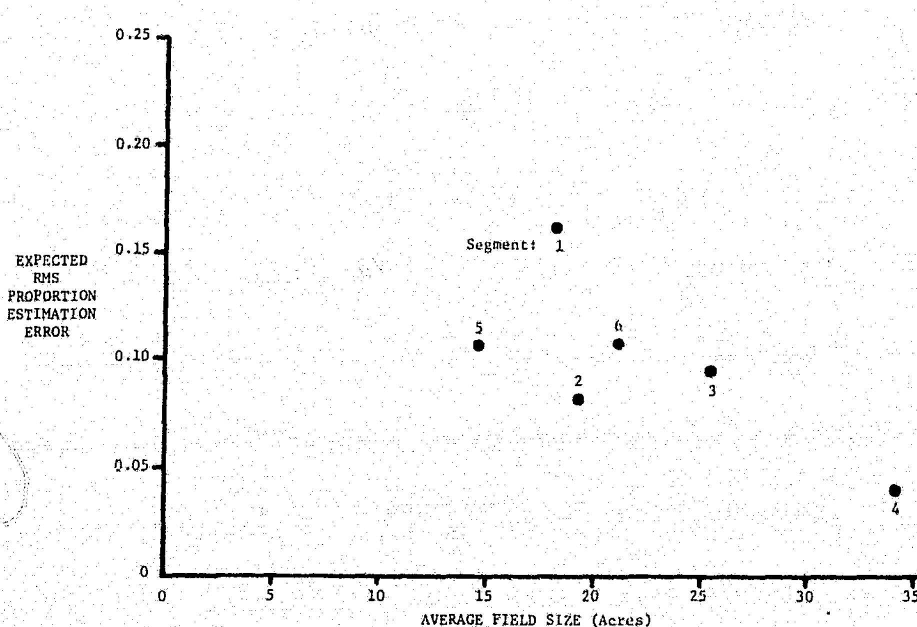


FIGURE 2. EFFECT OF FIELD SIZE (NO. BOUNDARY PIXELS) ON PROPORTION ERROR



FIRST RESULTS FROM THE CROP IDENTIFICATION TECHNOLOGY ASSESSMENT FOR REMOTE SENSING (CITARS)*

F. G. Hall
National Aeronautics and Space Administration
Johnson Space Center
Houston, Texas

M. E. Bauer
Laboratory for the Application of Remote Sensing
Purdue University
West Lafayette, Indiana

W. A. Malila
Environmental Research Institute of Michigan
Ann Arbor, Michigan

ABSTRACT

The CITARS task design, including objectives, analysis methodology and experimental procedures, is described and first results from this effort are presented. The extensive ground truth data set acquired for the CITARS task is described and discussed in some detail. Results of the accuracy tests for the photo interpretative CITARS ground truth are given. Results of the assessment of the ERTS MSS data for cloud cover and electronic quality are presented. Some results of the geometric correction and registration of the time sequential CITARS ERTS data are given. Finally, the field boundary selection problem is addressed and the results of the use of new technology for boundary selection are presented.

1. INTRODUCTION AND OBJECTIVES

In 1973, the Earth Observations Division (EOD) of the Johnson Space Center (JSC), the Environmental Research Institute of Michigan (ERIM), the Laboratory for the Application of Remote Sensing of Purdue University (Purdue/LARS), and the Agricultural Stabilization and Conservation Service (ASCS) of the U.S. Department of Agriculture (USDA), under took a joint task to quantify the crop identification performance, resulting from the remote identification of corn, soybeans and wheat using automatic data processing (ADP) techniques developed at ERIM, LARS, and EOD. These ADP techniques are automatic in the sense that subjective human interactions with the classification algorithms were minimized by the specification of the steps required for an analyst to convert a multispectral data tape to a classification result. The crop identification performances resulting from several basic types of ADP techniques are to be compared and examined for significant differences: The multispectral data to be analyzed, consists of ERTS-1 data acquired over each of six 5 x 20 mile segments in Indiana and Illinois at six periods from early June through early September 1973. Crop identification and other information was gathered by the ASCS in each segment each 18 days coincident with an ERTS overpass.

The ADP techniques are to be evaluated on this data set in two basic remote sensing situations: (1) Crop signatures for classifier training will be obtained within the same segment in which crops are recognized by the classifier (local recognition). (2) Crop signatures for classifier training will be obtained from a different segment in which crops are recognized (non-local recognition).

Once the crop identification performance is established for each of the ADP techniques for local and non-local recognition, differences in the performances of these techniques will be established for differences in geographic location, time of the year, etc.

* Proceedings of the Ninth International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, April 15-19, 1974.

The CITARS task was designed to quantitatively answer the following questions:

- o How does corn, soybeans, and wheat identification vary with time during the growing season?
- o How does the crop identification performance (CIP) vary among different geographic locations having different soils, weather, management practices, crop distributions, and field sizes?
- o Can statistics acquired from one time or location be used to identify crops at other locations and/or times?
- o How much variation in CIP is observed among different data analysis techniques?
- o Does use of multi-temporal data increase CIP?
- o Does use of radiometric preprocessing extend the use of training statistics and/or increase CIP?
- o How much variation in CIP results from varying the selection of training sets?
- o Does rotation or registration of ERTS data affect classification performance?

2. ANALYSIS METHODOLOGY

To establish and compare the capabilities discussed in the Introduction, an experiment was designed to: (1) accurately estimate the crop identification performance (CIP) and (2) determine whether differences in the CIP's for the various conditions are significant.

Each of the CIP's are established as a result of a specific "treatment" combination; such a treatment combination is characterized by several factors. These factors are: (a) ADP technique; (b) data acquisition period; (c) location; and (d) training-recognition method.

Each of these factors can, in turn, be characterized by levels. The levels of factor (a) are different ADP techniques to be assessed; the levels of factor (b)* are the six data acquisition periods from June through September 1973; the levels of factor (c) are the six test sites in Indiana and Illinois. There are many possible levels in factor (d) but they can be characterized for the present by (1) local recognition and (2) non-local recognition.

Each treatment combination will have an associated CIP which will be quantified in three ways: (1) a classification performance matrix from which the errors of omission and commission for "non-boundary" pixels can be determined and (2) a proportion classification error vector and (3) a proportion error vector corrected for bias.

The classification performance matrix for "non-boundary" pixels will be established by comparing the ADP classification with the ground and photo interpretive identifications of about 12,800 acres within each data segment. Test field boundaries will be established on the digital data. To insure that only non-boundary pixels are used in training and classification, the boundaries will be selected such that no agricultural field boundary elements or field inhomogenieties are contained within the test field boundaries. The probability for correct classification for each of corn, soybeans, wheat and "other" will be defined, for a particular test field set, as the frequency with which test field pixels of a particular class are correctly classified. The error of commission between class i and class j will be defined as the frequency with which an ADP identification of class i was determined from ground truth to actually have been a pixel from class j. For a four class data set this procedure will define a 4 x 4 error matrix.

The proportion classification error vector will be established by comparing the proportions of corn, soybeans, wheat and other as determined from the ADP technique to those proportions determined from ground truth. To establish the ground truth, twenty agricultural quarter sections in each segment were visited each 18 days by ASCS personnel for crop type identification. In addition, twenty additional agricultural sections (one mile square) were photo interpreted to establish crop identification.

*The levels in factor (b) will differ for the multitemporal ADP technique, i.e., if data from three passes are used for the analysis then there are ten possible ways to combine the six data acquisition periods.

The proportion of each crop type in the sections within each segment were established by mensuration of the photography. This results will be compared to the proportions determined by the ADP techniques to determine the ADP proportion error vector. In addition, several methods have been proposed to correct the remote sensing estimates of the crop proportions for bias. Each of these methods require an estimate of the bias, which is obtained by examining classification performance in fields or areas for which ground truth is available. These fields and areas will be called pilot fields or areas, to distinguish them from the test fields where the crop identification performance is to be established. The methods proposed for bias correction and the method for pilot field selection are more fully discussed later in this section.

Thus for each treatment a performance matrix and a proportion error vector can be estimated using the procedures described above.

These data, once computed, form a basis for comparison of the performances of the techniques under the various conditions. These comparisons will be made using standard statistical tests. The primary statistical test to be applied is the analysis of variance. The objective of such a procedure will be to test the hypothesis, that the classification performance for two or more different treatments (or combinations of treatments) are different. An example of a hypothesis to be tested is, "There are no significant differences in crop identification performance among test sites." To test this hypothesis, the ratio of variation among test sites is compared to the variation within test sites. This is referenced to as the calculated "F" which is the ratio of the treatment mean square (among) to the error mean square (within). If the calculated F is greater than a tabulated F based on the known distribution of the variance ratio under the null hypothesis, then one would reject the null hypothesis and accept the alternate hypothesis that performances are different for different locations. Similar hypotheses can be formulated for each factor. The comparisons of interest to the FY74 task discussed in the Introduction can be formulated into hypotheses and tested in the manner described above.

To use analysis of variance a measure of error must be available. This is obtained from replication which is readily available in a factorial experiment. For example, a mathematical model assumed is:

$$x_{ij} = \mu + \tau_i + \varepsilon_{ij} \ , \ i = 1, 2, \ldots, k, \ j = 1, 2, \ldots, n$$

which states that any observed value x_{ij} is equal to the overall mean μ for all populations, plus the deviation τ_i of the i-TH population mean μ from the overall mean, plus a random deviation from the mean of the i-TH population. In other words, if μ_i is the mean of the i-TH population, then

$$\mu = \text{Sum of } \mu_i/K$$

$$\tau_i = \mu_i - \mu$$

$$\varepsilon_{ij} = x_{ij} - \mu_i = x_{ij} - \mu - \tau_i$$

for this model it is assumed that

1. μ is an unknown parameter.

2. The τ_i are unknown constants or parameters.

3. The ε_{ij} are normally and independently distributed with mean zero and variance σ^2.

With estimates of the population means and variance σ^2 it is also possible to estimate the magnitude of treatment effects and to calculate confidence intervals.

3. EXPERIMENTAL PROCEDURES

3.1 TEST SITE SELECTION

The test sites were chosen over as large a geographic area as possible, within the resources of the project, in order to include a wide variety of conditions. Even in the Corn Belt there is a great deal of variation in soils, weather, cultural practices, crop distribution, etc. All of these factors are related to geographic location. The best measure of the effects of these factors, then, can be obtained by including as many test sites as possible over as large an area as possible. This also increases the probability of obtaining cloud-free ERTS data.

Test sites were selected within the four overlap zones of the five ERTS passes over Indiana and Illinois. These areas shown in Figure 1 include some of the different conditions which could be expected to be encountered in the Corn Belt.

3.2 SELECTION OF SEGMENTS AND SECTIONS

Segments, five by twenty miles, were chosen at random within each of the six selected counties. These segments were oriented such that the twenty mile length was oriented north-south. This segment size was chosen to give a limited area for field visits and yet an adequately large area for a representative sample of agriculture within a county.

Within each segment, 20 sections and 20 quarter sections were chosen at random in a manner such that the selected quarter sections were spatially disjointed from the sections selected.

3.3 FIELD OBSERVATIONS FOR CROP TYPE IDENTIFICATION

The ASCS of the USDA visited, each 18 days, the 20 quarter sections within each segment and examined each field in the quarter sections for crop type identification, other agricultural parameters shown in Figure 2. Atmospheric optical depth (related to visibility) at several locations, using tripod mounted solar spectrophotometers provided by JSC and subjective assessments of cloud cover weather and haziness during the ERTS overpass were also recorded by ASCS personnel.

3.4 PHOTO INTERPRETATION FOR CROP TYPE IDENTIFICATION

A more accurate estimate of the crop identification performance for each ADP technique can be obtained if a larger field sample is available from each segment. Thus, the field observation data was supplemented by photo interpretation of the 20 additional sections chosen in each segment.

The photo interpretation effort used large scale color IR (scale) photography acquired at five times during the growing season, and large scale metric photography acquired at two times, to establish proportions of ground cover classes and other agricultural parameters within each of the 20 sections in each segment.

A test procedure using ASCS visited quarter sections hidden in the photograph was devised to determine the accuracy of the crop identifications so obtained. The photo interpretation procedure was designed to obtain as accurate an identification as possible. When the PI test procedure indicated errors in the photo interpretation field identifications, the effects of these errors on the estimates of the ADP crop identification performance were assessed, once the source and nature of the photo interpretative errors were ascertained.

3.5 ERTS DATA PREPARATION

This section addresses those procedures required to reformat the ERTS MSS data, to locate in the data the sections and quarter sections, and to choose the test and training fields within these sections. These procedures have been designed to allow each institution to use common training and test field boundaries and duplicate ERTS data tapes. Such a procedure was followed to permit more meaningful performance comparisons and to eliminate needless duplication of tasks and resources at each institution.

ERTS bulk tapes were received at LARS for subsequent reformatting and field boundaries definition.

Of the two ERTS passes over each segment, the one acquired during minimum cloud cover was selected for local recognition. If cloud cover was equal for the two passes, the data acquired most temporally coincident with the ASCS field visit was chosen for local recognition processing.

3.5.1 GEOMETRIC CORRECTION AND REGISTRATION

ERTS data preparation for CITARS has consisted of (1) geometric correction, (2) multitemporal registration, and (3) section and field coordinate location. Geometric correction is performed to facilitate accurate location of section and field coordinates. Registration of the data from two or more ERTS passes over the same scene is required for multitemporal data analysis procedures and to reduce the number of times which section and field coordinates had to be located. With registered data the desired coordinates need to be found only once and the same coordinates are used for all data collected over the same area. Field and section coordinates are, of course, required for classifying the ERTS data and evaluating the results.

3.5.2 TEST, TRAINING, AND PILOT FIELD SELECTION

In addition to the training and test fields usually selected to train the classifier and to evaluate its performance, "pilot" fields were selected. The pilot fields will be used to determine

if a correction for bias in the classified crop proportion, resulting from classification errors, is feasible. Classification errors will be estimated in the pilot fields and based on these estimates, a correction will be applied to the test field classification results (See 3.8 FACTORIAL ANALYSES FOR PERFORMANCE COMPARISON for more details.)

For those analyses which require pilot fields, all fields from one-half of the 20 photo interpreted sections will be used for pilot fields and the remaining ten sections for test fields. For those analyses which require no pilot fields, all photo interpreted sections will be used as test fields.

Training fields from the ASCS quarter sections will be used to train the classifiers. All fields large enough to be accurately located in the scanner imagery will be available for training.

3.6 ADP TECHNIQUES FOR MSS DATA PROCESSING

The basic ADP techniques will be grouped into three divisions: (1) "Standard" techniques, (2) preprocessing techniques, and (3) multi-temporal techniques. The term "standard" ADP technique is used to mean either Gaussian maximum likelihood classifiers or a classifier employing a linear decision rule and classifies data which has not been radiometrically preprocessed and has not been acquired multi-temporally.

Each of these ADP techniques consists of a computer implemented software system and a method or procedure by which an analyst can convert multispectral data into ground cover class identification information on a pixel by pixel basis.

Since the crop identification performance of ADP techniques can be sensitive to the manner in which the classifier is trained, the types of MSS data input (e.g., preprocessed, multi-temporal, etc.) which spectral bands are used for recognition, etc. A quantitative evaluation and subsequent comparison of the crop ID performances of such techniques will be most meaningful if the procedures used to obtain the classification results are well defined and repeatable.

Most of the existing procedures currently developed for the use of the very generalized analysis algorithms, require decisions on the part of the analyst which can significantly affect the classification performance obtained. For the purposes of this assessment, the analyst factor will be minimized as much as possible in order to permit an evaluation of the automated techniques.

A necessary requirement for a small variance in the classification repeatability of an ADP technique is that the procedure for using such a technique be sufficiently well defined so that an analyst can follow the procedure without deviation; thus, each of the ADP techniques evaluated in this task will be documented in detail, and the documented procedures will be rigidly adhered to.

3.6.1 LARS ADP TECHNIQUES

The analysis techniques to be used by LARS utilize the LARSYS Version 3 multispectral data analysis system. Its theoretical basis and details of the algorithm implementation are described in references 1 and 2, respectively. A complete description of the analysis procedures is contained in reference 3. The procedures are designed to provide repeatable results, i.e., variation due to analysts is minimized. Briefly, the analysis procedures consist of:

1. Class Definition and Refinement. Four major classes, corn, soybeans, wheat (for selected missions) and all "other" ground covers are defined. These major classes are divided into subclasses where spectral variability within a class is so great as to result in multimodal probability distributions for that class. Clustering is used to isolate the subclasses. For clustering all four ERTS bands are used. A systematic method (see reference 3) which minimizes the total number of subclasses produced while ensuring that multimodal class distributions are avoided is used for interpreting information on the separability of subclasses.

2. Classification. Each data set is analyzed using two versions of the maximum likelihood classification algorithm. Gaussian probability density functions are assumed for both procedures. The first classification method is the maximum likelihood classification rule assuming equal prior probabilities for all classes and subclasses. This is the rule which has been in common usage for remote sensing data analysis for some time.

The second method uses "class weights" proportional to the class prior probabilities. This approach is more nearly optimal given that the Bayesian error criterion (minimum expected error) is preferred. Class weights may be based on any reasonably reliable source of information. In CITARS the weights are computed from county acreage estimates made by the USDA the previous year. Subclass weights are simply the number of points in each subclass divided by the total number of points in the class.

3. Results Display and Tabulation. The results of the classifications are displayed using a discriminate threshold of 0.1%. This light threshold eliminates only those data points very much different from the major class characterizations. Thresholded points are counted in the "other" category. A computer program is used to generate results tabulations, in both printed and punch-card form, for training fields, test fields, and test sections.

3.6.2 ERIM ADP TECHNIQUES

The digital data processing and analysis procedures defined by ERIM for use in the CITARS study reflect our concern for the calculational efficiency of recognition processors and the need for extending recognition signatures from training areas to other geographic locations and/or observation conditions, as well as the CITARS requirement for minimizing the need for analyst judgment. A brief summary of the procedures is as follows:

3.6.2.1 Training

The training of the processor, that is, the establishment of class signatures for recognition, is a crucial step in multispectral data processing. Although multimodal signatures are frequently employed, the use of one signature per major class was selected for CITARS processing because of simplicity, processing efficiency, and the fact that a combination of individual field signatures can result in a single signature that encompasses more of the variability of the class than is represented by a multimodal signature. An objective, reproducible procedure, based on a χ^2 test, was devised to reject anomalous "outlier" fields before the formation of a combined signature, so as to develop signatures representative of healthy crops at a reasonable stage of maturity for the time of season. Signatures for classes other than the major ones are to be included only if they are found to be confused with the major crops on preliminary recognition runs over training data.

3.6.2.2 Recognition Without Preprocessing

Two types of decision algorithms are being used, a linear rule and a more conventional quadratic rule. The linear decision rule was chosen because it requires substantially less computer time for recognition calculations, has been used successfully in many applications at ERIM, and has been found to provide comparable recognition accuracy in previous tests (reference 4). Use of the quadratic rule will permit another, comprehensive comparison of the two rules. Both rules apply a threshold test (0.001 probability of rejection) based on a quadratic calculation for the signature of the "winning" class; points failing the test will be classified as being other than the major crops considered.

3.6.2.3 Recognition With Signature Extension Preprocessing

Changes in atmospheric and other local conditions can cause changes in the signal levels received at the scanner for different areas and at different times. The region of signature applicability can be extended beyond the region used for training by employing signature-extension preprocessing techniques. Non-local recognition denotes recognition performed on segments other than those from which signatures were extracted. Non-local recognition will be carried out once before and once after preprocessing corrections for signature extension have been applied (for both linear and quadratic decision rules). Several promising preprocessing methods have been developed (references 5 and 6) and are being tested on ERTS data at ERIM (reference 7). Only one method has been identified to date for use on the CITARS project — a mean-level adjustment procedure. The mean-level adjustment is derived from an average over diverse ground covers within the "local" signature extraction segment and a comparable average within the "non-local" segment to be classified.

3.6.2.4 Results Summarization

The results obtained with each procedure will be summarized in a standardized form for subsequent analyses of variance. Separate summarizations will be made for field-center pixels and for entire test sections.

3.6.3 EOD ADP TECHNIQUES

EOD will evaluate two techniques. One technique for single pass data and another for multi-temporal MSS data.

For single pass data the EOD will utilize the ISOCLS (reference 8) clustering algorithm implemented at JSC to generate the class and subclass statistics and the Gaussian maximum likelihood classifier on the Earth Resources Interactive Processing System (ERIPS) (reference 9).

The training fields for corn, soybeans and wheat will be submitted to independent runs on the Earth Resources Interactive Processing System (ERIPS) using the ISOCLS implemented clustering routine

to generate class and, if necessary, subclass statistics, e.g., corn 1, corn 2, corn 3, etc. The training fields for "Other" will then be submitted to the same clustering scheme to generate class and subclass statistics for all "other".

The training fields, test fields and "test" sections will then be classified using the Gaussian maximum likelihood classification algorithm on ERIPS and the statistics as previously generated with the clustering process.

For multi-temporal data ISOCLAS will be used to separate spectral classes. A linear combination of features will be selected using and EOD algorithm (reference 10) and classification will be similar to the uni-temporal technique.

3.6 PERFORMANCE COMPARISONS

The basic questions proposed in the objectives will be answered by a series of analyses of variance. There will be two basic quantities which will be used to characterize the crop identification performance of the ADP techniques. These are (1) e_{ij}, the estimated probability of classifying a non-boundary pixel from class i as class j and (2) $\hat{p}_i - p_i$, the estimated proportion of class i minus the true proportion of class i.

In order to compute the e_{ij} from the ADP results, pixels which correspond to ground cover classes i and j must be precisely located with respect to the ground truthed areas. Test fields will be chosen to exclude agricultural field boundaries within pixels and also to exclude known inhomogeneities in the field, such as flooded areas, etc. As will be discussed later, accomplishing this task was difficult and required the development of new technology.

Since in an actual remote sensing situation, the classification error resulting from pixels containing agricultural field boundaries (boundary pixels) and the error resulting from field inhomogeneities may represent a large part of the error, some method is required to estimate these errors. The use of e_{ij} to accomplish this was decided to be impractical because of the difficulty in locating the pixels containing field boundaries. Thus it was decided to use the proportion estimate to characterize this error. $\hat{p}_i$ will be computed for "pure test pixels" as well as for the agricultural sections and the differences in the resulting proportion error vectors will be used to estimate the error contribution resulting from boundary pixels and field inhomogeneities.

In addition to the performance quantities discussed above, some attempt will be made to correct the proportion estimates for statistical bias which is expected to result from misclassification. Two methods have been proposed for accomplishing this and the corrected $\hat{p}_i$ using each method (described below) will be compared to the $\hat{p}_i$ as determined from the photo interpretation to determine if either method improves the proportion estimates.

3.7 FACTORIAL ANALYSES FOR PERFORMANCE COMPARISONS

For performance comparisons several dependent variables will be calculated for each of the 20 test areas per segment. The quantities e_{ij} will be estimated as discussed previously.

The proportion estimates $\hat{p}_i$ will be computed in one of three ways:

1. $\hat{p}_i = n_i/N$
2. $\hat{p}_i = \beta_i \, n_i/N$
3. $\hat{p}_i = \frac{1}{N} \; E^{-1}(n)$

where n_i = number of pixels classified as i.

N = total number of pixels in area to be classified.

β_i = regression coefficient obtained by comparing n_i/N with the true proportion p_i for pilot data

E = matrix of e_{ij}'s obtained from pilot data.

n = vector of n_i's.

Note that methods two and three require the use of "pilot" data, i.e., additional ground truth used to obtain estimates of E or β_i.

Once a dependent variable is decided upon a typical analysis will consist of (a) obtaining cell means of the dependent variable over various combinations of factors, (b) performing an analysis of variance.

4. FIRST RESULTS FROM CITARS

CITARS as designed and described above, was originally scheduled to begin in early June 1973 and to be complete by October 1974 with a majority of the actual ERTS data processing complete by April 1, 1974; However, as of April 1, 1974, CITARS is approximately 90 days behind schedule. This slip resulted primarily from difficulties encountered in field boundary coordinate location in the ERTS imagery (to be discussed momentarily).

Complete as of April 1 is the acquisition of ground truth by ASCS, the aircraft acquisition of large scale color IR photography, the interpretation of that photography for supplemental ground truth, data quality evaluation of the ERTS data, and the geometric correction and registration of that data. To be completed are Test, Training and pilot field coordinate determination, ADP processing of the data and the subsequent performance comparisons. The remainder of this paper will be devoted to a discussion of the completed portions of CITARS.

4.1 DATA ACQUISITION

Data acquisition consisted of three major efforts; the periodic visitation of each segment by ASCS personnel for crop type identification, the periodic acquisition of large scale aircraft photography and ERTS data acquisition. Each of these tasks have been completed and with some exceptions, each effort has provided data adequate for the accomplishment of the CITARS objectives.

4.1.1 GROUND OBSERVATIONS

ASCS personnel completed all field visits to all segments. In addition, they made additional visits in the Fayette County, Illinois segment to determine additional training and test fields for wheat which were required as a result of late and incomplete aircraft coverage of Fayette County in early June.

Table 1 summarizes for each county the amount of acreage, the total number of fields by class, and the average field size for the fields in the 20 ASCS June identified quarter sections. In addition to these and the other ag data discussed in section 3.3, ASCS personnel successfully used the solar photometers to record atmospheric optical depths on successive ERTS orbits over the segments. For some of these segments, there are considerable differences in the atmospheric state from one day to the next. Thus by training on a segment and classifying it on the subsequent ERTS orbit, the effect of atmospheric differences on crop identification performance can be evaluated.

4.1.2 AERIAL PHOTOGRAPHY

Aerial photography was used for field annotation, extension of ASCS ground truth via photo interpretation and mensuration of field acreage. For accurate photointerpretation, large scale color infrared photos were specified. This photography was acquired each 18 days from June through October by the Bendix Queen-Aire Aircraft at 4 km altitude using a Fairchild 224 camera. For accurate mensuration of fields, data was acquired in July and August with a Zeiss metric camera flown in the ERIM C-46 at about 4 km altitude. Base photography for the annotation of ground truth was acquired with an RC-8 camera carried at 18 km by the NASA RB57F.

Camera problems, excessive cloud cover and incorrect flight lines rendered some of the test and training sections in the Bendix photography unuseable. Some sections were also incompletely covered by the C-46. However, combining the photography from these two sources proved adequate for the CITARS requirements.

4.1.3 PHOTOINTERPRETATION

The photointerpretation team at EOD has completed the tasks of determining crop identification, areal proportion of each crop, row direction and width and any field anomalies for each of twenty agricultural sections in all CITARS segments. This job which took about 6 months from the acquisition of the first photography was completed within three weeks of the originally projected schedule. About 18 man months was expended in the effort. The photointerpreters, using large scale color IR photography acquired 6 times during the growing season, identified 23 or 24 agricultural sections in each segment. The photo interpreters trained in 16 or 17 of the ASCS quarter sections. Three or four of the remaining quarter sections were withheld from the photointerpreters, but were included in the full sections to be interpreted so that a comparison of their results to the ASCS identifications could be made. The photointerpreters did not know which sections contained the test quarter sections.

Comparisons of the ASCS crop identifications to the PI identifications have been completed in 3 of the 6 CITARS segments. In these counties the percent agreement with the ASCS identifications was 94% for soybeans, 96% for corn and 16% for wheat. These are percentages of fields with a sample of 32 soybean fields, 27 corn fields and 6 wheat fields. Of the 2 soybean fields misidentified as other, one field was planted 2 weeks late and the other one 6 weeks late. The corn field misclassification may be a result of mistaken field labeling. This problem is as yet unresolved. The wheat problem is a different matter since the first aircraft photography of June 28 was acquired during wheat harvest. Thus the photointerpreters had only mature (low IR reflectance) or harvested wheat imagery which proved inadequate for wheat recognition.

4.1.4 ERTS-1 MULTISPECTRAL SCANNER DATA

The ERTS-1 satellite passed over each of the six test segments twice (on successive days) during each 18-day period. Since there were six periods of interest, from early June to early September 1973 a total of 72 data sets were potentially available for processing and analysis. Cloud cover problems were identified by reference to the ERTS data catalog and visual inspection of imagery where available. Of the 72 possibilities, cloud cover was severe enough (20-30%) on 41 sets to cause their rejection outright, no data were collected for two sets, and several others were eliminated for other reasons. A total of 23 sets were selected for analysis and several of these were found to have cloud problems upon detailed examination. Thus, roughly 60% of the data sets were eliminated because of excessive cloud cover. Table 2 summarizes the cloud cover statistics.

A majority of the selected data were of good quality. However a few problems were observed which are affecting data analysis procedures and/or results: (a) occasional erratic data throughout individual scan lines or portions of lines, (b) detector-to-detector* differences among the mean values obtained from the six detector channels that comprise each spectral band as averaged over a large sample of the data and (c) differences in the variances observed from the detector channels over the same data sample.

ERTS-1 data quality was assessed by several different methods. First, visual inspection was made for each spectral band on a digital display to determine the presence of any lines of bad data through one of the 5 x 20 mile segments. More than half (14) of the 23 sets had no bad lines, and the worst were one with 8, two with 19, and one with 40 bad lines. Next, histograms and sample statistics (mean and standard deviation) were computed for samples of the data — every line and every 30th point for all cases, every line and every sixth point for many, and every line and every point for a few. These statistics were calculated separately for each detector in each spectral band; unrotated ERTS data were utilized for these tests.

One would expect some variation between values in the various detectors, because each is calibrated separately. To evaluate the degree of similarity between these statistics, a mean, m_μ, of the six detector means was calculated for each spectral band, as well as the sample standard deviation, s_μ, of the individual values from that overall mean. The ratio, s_μ/m_μ, was computed for each data set and a histogram of these values is presented in Figure 3. All values lie below 3%, except for one of 9.3%, which corresponds to the data set with 40 lines of bad data. No clear relationship could be found between the number of bad lines and s_μ/m_μ values below 3%. The number of good lines present was sufficient to mask effects of a few bad lines, and channel-to-channel variations existed in all data.

Similarly, the detector value standard deviations were analyzed and a histogram of s_σ/m_σ is presented in Figure 3. In this instance, the s/m values exhibit considerably more spread than they do for the detector means. Most values are $\leq$5%, but they scatter up to 24%, with an extreme of 87% for the 40-bad-line case. Here again, except for extreme values there was no apparent direct correlation between s_σ/m_σ and the number of bad lines present.

The question now arises regarding the analysis of data exhibiting problems associated with clouds, lines of bad data, and channel-to-channel variations. Test section, test fields, and training fields affected by clouds and/or bad lines were determined by inspection and eliminated from the analysis. The one data set with 40 bad lines is being analyzed in three bands only, since all bad lines were in the same band. One could transform all data values to equalize the means and/or variances in each set of detector channels (omitting bad-line values) and perhaps effect some improvement in recognition results. However, geometric correction and spatial registration operations were being applied to these data sets in parallel with the data quality evaluations, so it was decided to start again and carry out radiometric correction procedures only if poor recognition performance were obtained and appeared to be attributable to such differences.

*Detector is used here to denote the entire signal channel from individual detector element to CCT.

4.2 ERTS MSS DATA PREPARATION

4.2.1 GEOMETRIC CORRECTION

The digital form of the ERTS data (CCT's) contains several geometric distortions. These distortions include: scale differential, altitude and attitude variations, earth rotation skew, orbit velocity change, scan time skew, nonlinear scan sweep, scan angle error, and frame rotation. The major errors are the scale and skew errors. Also, rotation to North-orientation is highly desirable. A two-step process, developed by LARS, to geometrically correct ERTS data over small areas has been applied to all data for CITARS (reference 11).

Briefly, the procedure uses four linear transformations to correct or adjust for horizontal and vertical scale differences, rotation, skew due to earth rotation and output scale factor. The process assigns radiance values in a rescaled, rotated, and deskewed coordinate system using data from the existing grid, i.e., the raw ERTS data. Because a fixed grid output device is used (i.e., line printer or digital display screen) some interpolation is required to produce new samples. The point nearest the desired sample location is used to represent the value at the desired location ("Nearest Neighbor Rule").

The procedures are fully described in reference 11. The output form used for CITARS is such that when the data is printed on an 8 line per inch, 10 column per inch computer line printer the resulting scale is approximately 1:24,000 and the image is North-oriented (Figure 4). Comparisons made using topographic maps indicate about a 1 to 2% scale error.

4.2.2 SPATIAL REGISTRATION OF TIME SEQUENTIAL ERTS DATA

Registration of multiple images of the same scene is accomplished through use of the LARS image registration system described in reference 12. The overlay processing operation consists of two basic operations: (1) image correlation and (2) overlay transformation. Many factors exist which prevent exact overlay of the images, thus this operation is approximate. Two major errors are: (1) It is unlikely that the samples from one time were imaged from exactly the same spot as samples from a later satellite pass; thus, in general, no data exists which exactly overlays for both times even if no other errors were present; and (2) Due to changes in the scene and other "noise" sources the two images cannot be exactly correlated or matched. The overlay procedure used consists of the following:

1. Initial checkpoints or matching points are manually selected in the two images to be overlayed using a digital display screen.

2. A two dimensional least squares quadratic polynomial is generated to represent the difference in position of points in the two images.

3. A block image cross correlator is employed to find the remaining image displacement at the nodes of a uniform grid using the approximate overlay polynomial generated in reference 12.

4. A new overlay polynomial is generated from the correlator produced set of checkpoints and used to actually overlay the images. The two images are combined onto one data tape and a new data set is formed having M+N channels where M is the number of channels from image A and N is the number of channels from image B.

5. The overlay data tape is inspected to check image quality and overlay quality. Precise evaluation of overlay accuracy is difficult. A measure of registration error is obtained from the residual of the least squares polynomial; this statistic averages 0.5 of an image sample, RMS (Table 3). However, crop identification performance obtained when using field coordinates selected from a base period will be compared to performance when using coordinates located in data which has not been registered.

4.7.3 SECTION AND FIELD COORDINATE LOCATION

Locating section and field coordinates in the ERTS data has been a major task preparatory to classifying the ERTS data. This task was first attempted using a manual method for location of fields in ERTS data displayed as single-band gray-scale line printer maps (reference 13). This required that field boundaries be easily distinguished in the imagery. In cases where there was minimal spectral contrast among crop fields, non-supervised classifications have been performed to produce an enhanced image. Whether using gray-scale or computer enhanced images, reasonably large fields are required to assure that pixels are selected from within the field boundaries.

With the CITARS data, there was little contrast among fields of interest, since the first data was collected early in the growing season (June 8-12). For instance, at this time of year corn and

soybeans were only a few inches tall and the spectral response was primarily from the soil. And, roads were not as visible in the imagery as they generally are in data collected later in the season. Also, many fields were small (≤ 20 acres). Therefore, procedures for accurately locating fields, when individual fields could not be clearly seen in the imagery, were required to meet project requirements.

To improve the accuracy of the manual location method, ERTS images were geometrically corrected and and rescaled to produce a nominal 1:24,000 scale map on a line printer (reference 11). This product alone made the location of fields more precise and more rapid than it would have been on uncorrected data. Photo overlays were prepared with section and field boundaries outlined. The initial overlays, made from photography enlarged to a nominal scale of 1:24,000, were helpful, but not completely satisfactory due to distortions in the photo. Following this rectified photographs were produced at a scale of 1:24,000. This product could be manually overlaid to the line printer maps of the ERTS data.

After manually locating all field and section coordinates in the ERTS data the precision was still not adequate to meet the requirement of a maximum error of one pixel. Therefore, a previously developed, computer-assisted procedure was employed by ERIM to locate section corners and define ERTS data coordinates for sections (reference 14). A map transformation from Earth coordinates on a rectified aerial photograph to ERTS data coordinates was calculated for each segment using roughly 30 control points for each calculation. The control points were located visually in the rotated and geometrically corrected ERTS data and by coordinate digitization on the photograph. A map transformaton then was computed by the method of least squares; ERTS coordinates of the few control points with large residuals (>1 pixel) were checked and modified or deleted, as appropriate, and the transformation was recomputed. Next, the transformation was applied to all section corners of interest (whose locations on the photograph had been digitized at the same time as the control points) to find their fractional line and column coordinates in the ERTS data. Final standard errors of estimate (for control points) were less than 0.5 and typically between 0.2 and 0.4 ERTS pixels, i.e., 15 to 30 meters on the ground. The RMS error in digitizing the location of the individual points was on the order of three meters on the ground (errors of roughly 0.005 inch or less on a photograph at a scale of 1:24,000).

These section corner coordinates (calculated in fractional ERTS line and column coordinates) were used to locate field boundaries of individual fields within the sections. A major advantage of the procedure is that it preserves the relative positions of all points considered with an accuracy that cannot be matched manually. Another feature of the ERIM procedure was utilized to generate ERTS data coordinates for each outlined section. All pixels whose centers fell inside lines connecting the vertices (again, located by fractional coordinates) were automatically included on coordinate definition cards.

4.3 ADP ANALYSIS OF THE ERTS DATA

As a result of the difficulties encountered with the field boundary selection problem, the ADP analyses of the ERTS data and the subsequent performance comparisons as specified by the CITARS design plan, have not been completed. However, the preliminary analyses to date merit some brief discussion.

The most significant result to date, is the amount of training data acquired from the 20 quarter sections in each of the 5 x 20 mile segments.

To insure that the pattern classifiers were being trained only on "pure" and correctly identified ground classes, the CITARS task design specified that training data come from the quarter sections visited by ASCS personnel and that no pixel (ERTS data resolution element) which contained a boundary between different ground classes be used in the computation of training statistics. During the CITARS design phase, the amount of ground truth required for classifier training was estimated by assuming that ten times the number of channels used for classification would be required to train the classifier for each ground cover class. Thus based on 20 ground cover classes, and four channels, 800 "pure" or non-boundary pixels would be required for training. Other rough calculations (reference 15) indicated that no more than about one half of all acquired pixels would contain agricultural boundaries based on a preliminary estimate of a 20 acre average field size for Indiana and Illinois. Thus for ERTS pixels of 1.1 acres in size, roughly 1600 acres would be required to obtain the 800 "pure" training pixels. In addition, an equal additional amount of training data was required to form replicate training sets to determine the effect of training set selection on classifier performance. Thus ASCS was requested to visit and identify twenty 160 acre quarter sections to obtain 3200 acres for training purposes.

It is of interest to see how this design worked out in practice. In each of the CITARS segments, training field boundaries have been selected and final boundary adjustments are nearing completion. Based on the number of training pixels selected to date, Figure 5 plots the percent of the training acres actually selected as "pure" training pixels versus the average field size

in the ASCS quarter sections. Except for the anomoly in Livingston County (yet to be explained) the shape of this curve is as one would expect. However, the percent usable pixels are much smaller than the early estimates of 50%. This is a result of a subsequent design change in CITARS to include a row of "guard" pixels between the agricultural field boundary and the training field boundary. This was done to increase the probability that only non-boundary pixels were chosen for training, but prevented the selection of pixels in many fields, especially ones less than ten acres (20-40% of the fields) and larger but narrow fields.

5. SUMMARY

The CITARS task was designed as carefully as possible to insure an objective and quantitative assessment of the crop identification performance of currently available classification algorithms and procedures.

The ADP procedures were written to minimize the amount of subjective human interaction. This was done to permit a quantitative and repeatable evaluation of classification techniques which could be automated for operational implementation.

For the resources available, a data set was designed to permit an objective evaluation of these techniques over a wide as possible range of agricultural and climatological conditions. An extensive data set has been collected and the utmost care has gone into the preparation of this data set for ADP crop identification performance evaluation.

Based on the data set acquired and the stated objectives of CITARS, a factorial analysis has been designed to obtain the maximum amount of reliable information regarding classifier performance.

At this point in the progress of the CITARS task, the most major problem encountered was the selection of field boundaries in the ERTS data. This problem resulted from the requirement that no pixels which contained boundaries between different agricultural classes were included in the training or test data. This problem had to be resolved through the implementation of recently developed technology and has resulted in a 90 day delay.

At this point, the combination of ASCS field visits with interpretation of low altitude temporally acquired photography appears to be a relatively cost effective and accurate method for assumbling a large ground truth set with stringent design requirements.

Of the 72 possible ERTS acquired data sets roughly 60% of the data sets were unusable as a result of excessive cloud cover. Of the remaining sets the electronic data quality was acceptable for processing.

ACKNOWLEDGEMENTS

The authors wish to thank those at ERIM, LARS, and EOD, who devoted and are continuing to devote their time, talents and effort to the design, implementation and execution of the CITARS task. There are too many to list here, but their work should gain the recognition it deserves as CITARS results appear in the literature.

REFERENCES

1. P. H. Swain, 1972. "Pattern Recognition: A Basis for Remote Sensing Data Analysis", LARS Information Note 111572, Laboratory for Applications of Remote Sensing, Purdue University, W. Lafayette, Indiana.

2. Laboratory for Applications of Remote Sensing Staff, 1973. "LARSYS Version 3 User's Manual." T. L. Phillips (ed.). Laboratory for Applications of Remote Sensing, Purdue University, W. Lafayette, Indiana.

3. B. J. Davis and P. H. Swain, 1974. "An Automated and Repeatable Data Analysis Procedure for Remote Sensing Applications." Proceedings of Ninth International Symposium on Remote Sensing of Environment. Ann Arobor, Michigan, April 15-19, 1974.

4. R. B. Crane and W. Richardson, "Rapid Processing of Multispectral Scanner Data Using Linear Techniques", Presented at the Conference on Earth Resources Observation and Information Analysis System, Tullahoma, Tennessee, 13-4 March 1972. Report 31650-107-S.

5. F. J. Kriegler, W. A. Malila, R. F. Nalepka, and W. Richardson, "Preprocessing Transformations and Their Effects on Multispectral Recognition", Proceedings of the Sixth International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, October 1969.

6. R. F. Nalepka and J. P. Morgenstern, "Signature Extension Techniques Applied to Multispectral Scanner Data", Proceedings of the Eighth International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, 2-6 October 1972.

7. W. A. Malila and R. F. Nalepka, "Advanced Processing and Information Extraction Techniques Applied to ERTS-1 MSS Date", Presented at the Third ERTS Symposium, Washington, D. C., December 10-13, 1973.

8. R. T. Minter, "Computer Program Documentation ISOCLS Iterative Self Organizing Clustering Program, Program C094", Prepared by Lockheed Electronics Company under contract NAS9-12200, October 1972.

9. Earth Resources Interactive Processing System (ERIPS) User's Guide, Volume I and II, Prepared by IBM under contract NAS9-996.

10. K. Baker, "A User's Guide to the University of Houston Linear Feature Selection Program (Program 0010)", NASA Working Document, October 12, 1973.

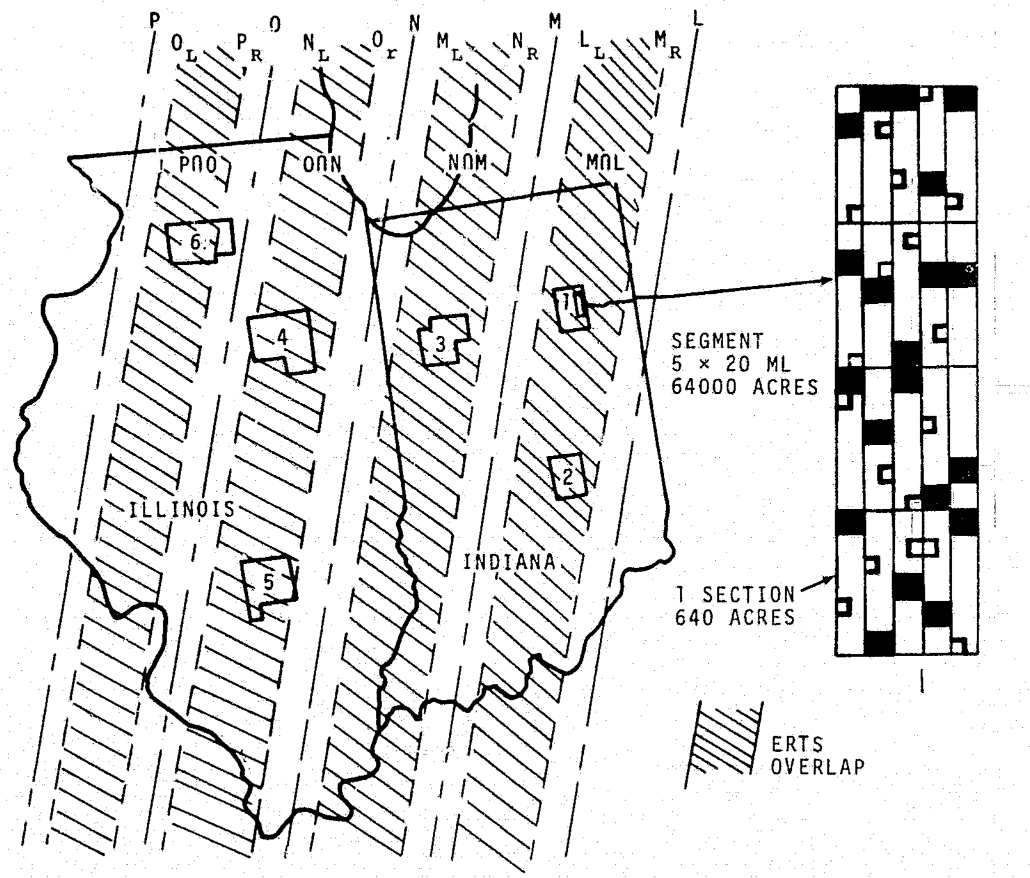
11. P. E. Anuta, 1973. "Geometric Correction of ERTS-1 Digital Multispectral Scanner Data", LARS Information Note 103073, Laboratory for Applications of Remote Sensing, Purdue University, W. Lafayette, Indiana.

12. P. E. Anuta, 1970. "Spatial Registration of Multispectral and Multitemporal Digital Imagery Using Fast Fourier Transform Techniques", IEEE Transactions on Geoscience Electronics, 8:353-368.

13. M. E. Bauer and J. E. Cipra, 1973. "Identification of Agricultural Crops by Computer Processing of ERTS MSS Data", Proceedings of Symposium of Significant Results Obtained from ERTS-1, New Carrollton, Maryland, March 5-9, 1973, pp. 205-212; and LARS Information Note 030173.

14. W. A. Malila, R. H. Hieber and A. P. McCleer, 1973. "Correlation of ERTS MSS Data and Earth Coordinate Systems", Proceedings of Conference on Machine Processing of Remotely Sensed Data, Purdue University, W. Lafayette, Indiana, October 16-18, 1973. ERIM No. 193300-18SA/J.

15. R. F. Nalepka and P. D. Hyde, "Estimating Crop Acreage from Space Simulated Multispectral Scanner Data", NASA Report NAS9-9784, p. 8, August 1973.



STUDY AREA COUNTIES

ILLINOIS	INDIANA
4. LIVINGSTON	1. HUNTINGTON
5. FAYETTE	2. SHELBY
6. LEE	3. WHITE

GROUND TRUTH

ASCS — 20 QUARTER SECTIONS (WHITE) EACH ERTS PASS

PHOTO INT. — 20 SECTIONS (BLACK) EACH ERTS PASS

FIGURE 1. CITARS ERTS DATA SET DESIGN

ASCS SUPPORT FOR CROP ID TECHNOLOGY ASSESSMENT
GROUND OBSERVATIONS SUMMARY

COUNTY ____________ DATE ____________ OBSERVER ____________

TOWNSHIP ____________ RANGE ____________ SECTION ______ (NW, NE, SE, SW) QUARTER

FIELD ID 1	ACRES (EST.) 2	CROP 3	HEIGHT 4	STAGE OF MATURITY 5	ROW DIRECTION 6	ROW WIDTH 7	GROUND COVER 8	OBSERVATIONS 9

ROW DIRECTION KEY-COLUMN 6
0 - NO ROWS
1 - N - S
2 - E - W
3 - NW - SE
4 - SW-NE
5 - CONTOUR

GROUND COVER KEY-COLUMN 8
0 - 0-5%
1 - 5-20%
2 - 20-50%
3 - 50-80%
4 - 80-100%

COMMENTS: ____________

OBSERVATION KEY
0 - NO COMMENT

EARLY IN CYCLE:
1-BARE SOIL
2-FRESHLY CULTIVATED
3-NEWLY SEEDED
4-VOLUNTEER
5-REPLANTED
6-REGROWTH

HOMOGENIETY:
11-THIN STAND
12-DROWNED SPOTS
13-SKIP ROW PATTERN
14-STRIP FARMING

LATE IN CYCLE:
21-HARVESTED
22-PARTHARVESTED
23-GRAZING
24-DEFOLIATED
25-WIND ROWED
26-CHOPPED FOR SILAGE
27-TILLED AFTER HARVEST

STRESS FACTORS:
31-DROUGHT DAMAGE
32-EXTREMELY WEEDY
33-PLANT DISEASE
34-NUTRIENT DEFICIENCY
35-HAIL DAMAGE
36-LODGING
37-INSECT DAMAGE

SURFACE MOISTURE:
41-DRY
42-MOIST
43-WET
44-STANDING WATER
45-IRRIGATING

OTHER:
51 ____________

FIGURE 2. ASCS GROUND OBSERVATION SUMMARY FORM FOR CROP IDENTIFICATION AND CONDITION

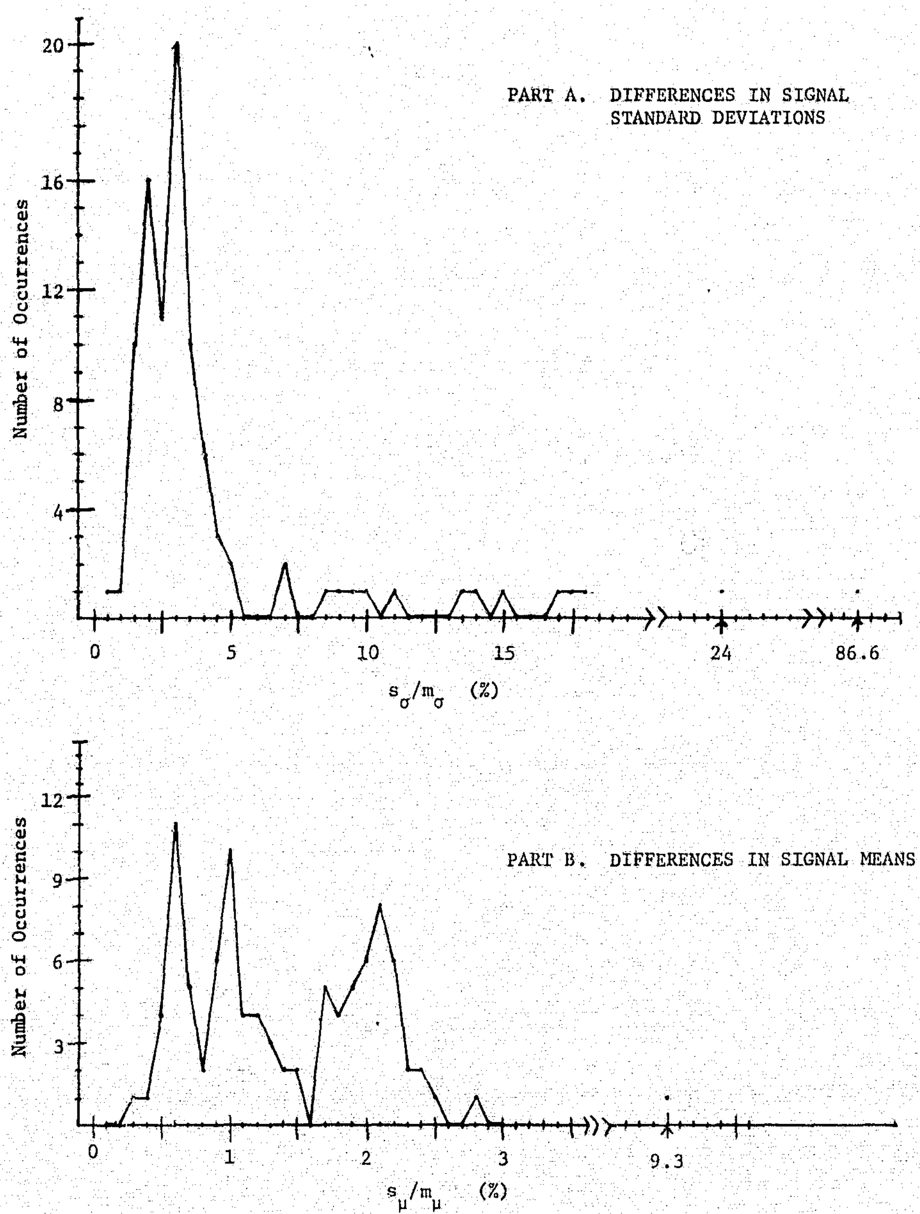
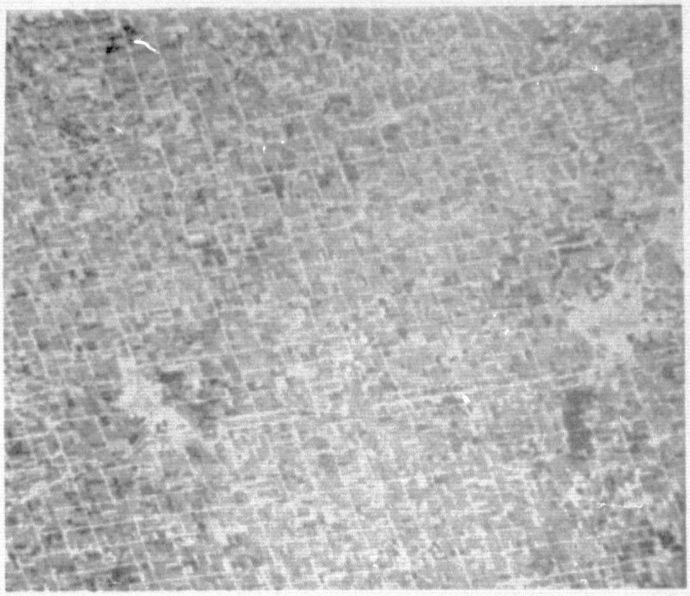


FIGURE 3. VARIABILITY OBSERVED BETWEEN THE SIX DETECTOR CHANNELS IN SINGLE BANDS OF ERTS DATA



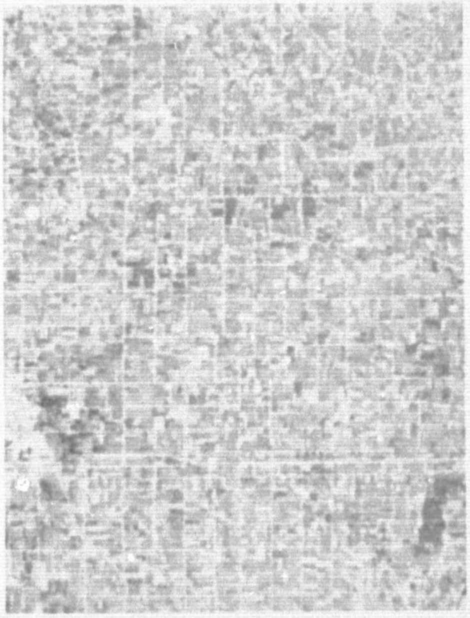


FIGURE 4. Comparison of original and geometrically corrected and rotated ERTS-1 MSS digital imagery. Upper image is the original. Lower image is skew and scale corrected and rotated to North. Scale is such that when this data is printed on an 8 line per inch, 10 column per inch computer lire printer, the resulting scale will be approximately 1:24,000.

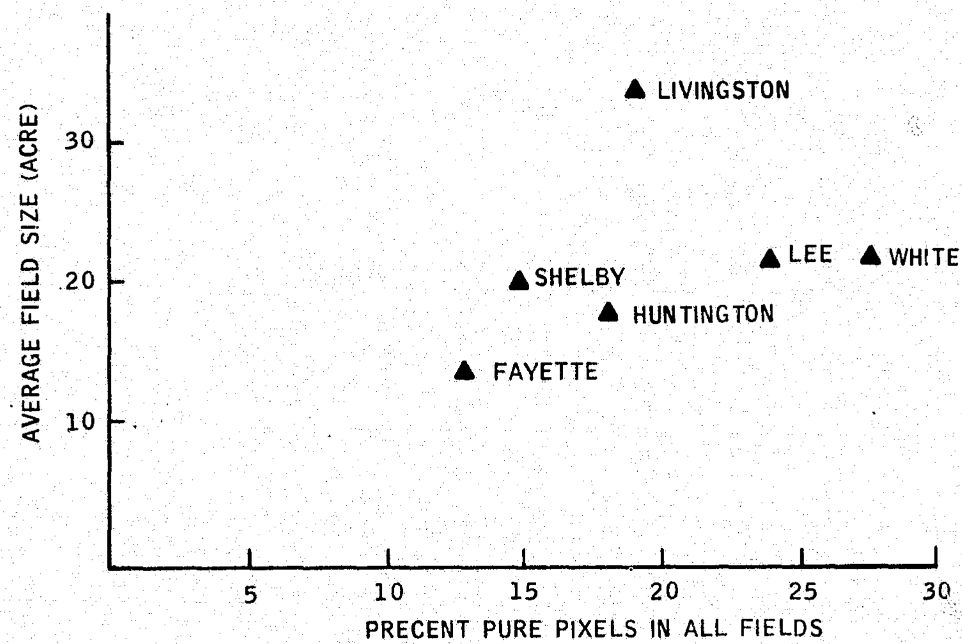


FIGURE 5. Percent of Non-boundary Pixels Selected for Classifier Training Versus the Average Field Size in the ASCS Quarter Sections

TABLE 1. SUMMARY OF ASCS IDENTIFIED QUARTER SECTIONS

COUNTY		CORN	SOY	WHEAT	OTHER	TOTAL
Lee	ACRES	1498	813	36	620	3550
	NO. FIELDS	42	31	2	34	160
	AVG. SIZE	35.6	26.2	18.0	18.2	22.1
Livingston	ACRES	1239	1073	39	569	2969
	NO. FIELDS	33	27	2	33	87
	AVG. SIZE	37.5	39.7	19.5	17.2	34.1
Fayette	ACRES	733	287	416	1358	3193
	NO. FIELDS	37	11	26	92	217
	AVG. SIZE	19.8	26.0	16.0	14.7	14.7
White	ACRES	1836	510	38	954	3753
	NO. FIELDS	42	13	2	41	146
	AVG. SIZE	43.7	39.2	19.0	23.3	25.7
Shelby	ACRES	1888	540	323	753	3648
	NO. FIELDS	71	24	15	61	189
	AVG. SIZE	26.5	22.5	21.5	12.3	19.3
Huntington	ACRES	831	618	63	986	2756
	NO. FIELDS	39	25	6	54	148
	AVG. SIZE	21.2	24.7	10.4	18.3	18.6

Table 2. Percent Cloud Cover over CITARS Test Sites During ERTS Passes.

Segment	Pass*	June 8-12	June 26-30	July 14-18	August 1-5	August 19-23	September 6-10
		Date					
Huntington Co., Indiana	1	A	X	X	X	X	X
	2	A	X	A	X	A	X
Shelby Co., Indiana	1	A	X	B	X	C	B
	2	X	X	X	X	X	X
White Co., Indiana	1	A	X	X	X	X	X
	2	X	X	X	X	A	C
Livingston Co., Illinois	1	A	X	B	B	C	X
	2	X	C	X	X	X	X
Fayette Co., Illinois	1	A	X	A	X	A	X
	2	A	A	A	X	X	X
Lee Co., Illinois	1	X	X	B	X	X	X
	2	X	X	A	A	X	X

Percent Cloud Cover

A 0 - 5
B 6 - 15
C 16 - 30
X 31 - 100

*Segments are located in overlap areas between ERTS passes on successive days

Table 3. RMS Error of Spatially Registered Multi-temporal ERTS Data.

Segment	Period	Number Checkpoints	RMS Error Lines	RMS Error Columns
Huntington	1	30	.44	.36
	3	27	.31	.39
	7	51	.30	.43
Shelby	3	23	.63	.38
	5	9	.27	.27
	6	43	.30	.47
	7	59	.31	.43
White	5	61	.32	.39
	6	16	.28	.14
Livingston	2	32	.31	.44
	3	9	.44	.37
	4	9	.16	.87
Fayette	1	41	.39	.33
	2	52	.37	.29
	3	53	.44	.34
	5	19	.57	.39
Lee	3	100	.34	.58
	4	84	.32	.41

Mr. Robert B. MacDonald/TF (1)
Dr. Andrew E. Potter/TF3 (8)
Dr. F. Hall/TF2 (1)
Mr. B. Erb/TF2 (1)
Mr. J. Dragg/TF4 (1)
Earth Resources Data Facility/TF12 (8)
J. Murphy/TF2 (1)

NASA/Johnson Space Center
Earth Observations Division
Houston, Texas 77058

Mrs. R. Elder/BB63 (1)
NASA/Johnson Space Center
Technical Support Procurement
Houston, Texas 77058

Mr. D. W. Mooneyhan (1)
Earth Resources Laboratory, GS
Mississippi Test Facility
Bay St. Lousi, Mississippi 39520

Dr. G. Thorley
EROS Data Center
USGS National Center, Reston, Virginia 22092

Dr. Larry Guseman (1)
Department of Math
Texas A&M University
College Station, Texas 77843

Mr. Eugene Davis/FA (1)
NASA/ Johnson Space Center
Computation & Flight Support
Houston, Texas 77058

Dr. Robert Miller (1)
U.S. Department of Agriculture
Agricultural Research Service
Washington, D.C. 20242

Dr. Craig Wiegand (1)
U.S. Department of Agriculture
Soil & Water Conservation Research Division
P.O. Box 267
Weslaco, Texas 78596

Dr. James R. Anderson (1)
Mail Stop 115
Geography Program
USGS National Center
Reston, Virginia 22092

Mr. Victor I. Myers (1)
Director, Remote Sensing Institute
South Dakota State University
Agriculture Engineering Building
Brookings, South Dakota 57006

Mr. Harvey K. Nelson (1)
U.S. Department of Interior
Fish & Wildlife Service
Bureau of Sport Fisheries & Wildlife
Northern Prairie Wildlife Research Center
Jamestown, North Dakota 58401

Dr. Richard Driscoll (1)
U.S. Department of Agriculture
Forest Service
240 W. Prospect Street
Fort Collins, Colorado 80521

Mr. A. E. Coker (1)
U.S. Department of Interior
Geological Resources Division
500 Zack Street
Tampa, Florida 33602

Mr. John M. DeNoyer
Director EROS Project
Department of Interior
1925 Newton Square, East
Reston, Virginia 22060

Mr. W. A. Fischer (1)
U.S. Department of Interior
Geological Survey
GSA Building, Room 5213
Washington, D.C. 20242

Mr. James Bettle
NASA Wallops
Wallops Station, Virginia 23337

Dr. J. Erickson (1)
Dr. R. Nalepka (1)
H. Horowitz (1)
Mr. R. Crane (1)

Environmental Research Institute of Michigan (ERIM)
P.O. Box 618
Ann Arbor, Michigan 48107

Dr. Raymond W. Fary (1)
U.S. Department of Interior
EROS Office
Washington, D.C. 20242

Mr. Charles Withington (1)
U.S. Department of Interior
Geological Survey
801 19th Street, N.W.
Washington, D.C. 20242

Mr. M. Deutsch (1)
U.S. Department of Interior
Geological Survey
801 19th Street, N.W.
Washington, D.C. 20242

Dr. Jules D. Friedman (1)
U.S. Geological Survey
801 19th Street, N.W., Room 1030
Washington, D.C. 20242

Dr. Harry W. Smedes (1)
U.S. Department of Interior
Geological Survey
Federal Center
Denver, Colorado 80225

Mr. Aaron L. Higer (1)
U.S. Department of Interior
Geological Survey
Water Resources Division
901 S. Miami Ave.
Miami, Florida 33130

Dr. Robert Colwell (1)
University of California
School of Forestry
Berkeley, California 94720

Dr. Charles E. Poulton
Earth Sat Corporation
Oregon State University
Corvallis, Oregon

Mr. William Hemphill (1)
U.S. Department of Interior
EROS Office
Washington, D.C. 20242

Mr. Leslie Dunn (1)
Chief of Technical Support
Western Environment Research Laboratories
Environmental Protection Agency
P.O. Box 15027
Las Vegas, Nevada 89114

Mr. William Howle
NASA/Langley Research
Mail Stop 470
Hampton, Virginia 23365

Mr. Kenneth Watson (1)
U.S. Geological Survey
Branch of Regional Geophysics
Denver Federal Center, Building 25
Denver, Colorado 80225

Mr. J. W. Sherman, III (1)
NAVOCEANO, Code 7001
Naval Research Laboratory
Washington, D.C. 20390

Mr. Kenneth Frick (1)
U.S. Department of Agriculture
Administrator
Agricultural Stabilization and Conservation Service
Washington, D.C.

Mr. R. C. Heller (1)
Pacific Southwest Forest & Range Experiment
Station
U.S. Forest Service
P.O. Box 245
Berkeley, California 94701

Dr. P. Weber (1)
Pacific Southwest Forest & Range Experiment
Station
U.S. Forest Service
P.O. Box 245
Berkeley, California 94701

Dr. Henry Decell (1)
Department of Math
University of Houston
Houston, Texas 77004

Dr. M. Stuart Lynn (1)
Institute for Computer Services and Applications
Rice University
Houston, Texas 77001

Mr. M. Kolipinski (1)
U.S. National Park Service
Western Regional Office
450 Golden Gate Avenue
San Francisco, California 94102

G. F. Hart/W. H. Wigton (2)
U.S. Department of Agriculture
Statistical Reporting Service
Room 4833 South Bldg.
Washington, D.C. 20250

Dr. Patrick L. Odell (1)
University of Texas at Dallas
Box 688
Richardson, Texas 75080

Mr. W. Nordberg, 620 (1)
Mr. W. Alford, 563 (1)

Goddard Spaceflight Center
National Aeronautics and Space Administration
Greenbelt, Maryland 20771

Dr. Herman Mark (1)
Lewis Research Center
National Aeronautics and Space Administration
21000 Brookpark Road
Cleveland, Ohio 44135

Mr. J. P. Claybourne/AA-STA (1)
John F. Kennedy Space Center
National Aeronautics and Space Administration
Kennedy Space Center, Florida

Mr. James L. Raper (1)
NASA/ Landgley
Mail Stop 214
Hampton, Virginia 23665

Dr. H. O. Hartley (1)
Texas A&M University
Institute of Statistics
College Station, Texas 77843

Dr. T. Bullion (1)
Texas Tech University
Department of Math
P.O. Box 4319
Lubbock, Texas 79409

Mr. James D. Nichols
ESL Incorporated, 495 Java Drive
Sunnyvale, CA 94086

Mr. J. O. Bennett (1)
EXXON Production Research Co.
P.O. Box 2189
Houston, Texas 77001

Mr. Andy Benson
Remote Sensing Lab, Room 260
University of California, Berkeley
Berkeley, California 94720

Mr. W. E. Kibler, Administrator (1)
U.S. Department of Agriculture
Statistical Reporting Service
Washington, D.C. 20250

Dr. James A. Smith (1)
Department of Watershed Sciences
Colorado State University
Fort Collins, Colorado 80521

Mr. R. Tokerud (1)
Lockheed Electronics Co.
16811 El Camino Real
Houston, Texas 77058

Dr. David Detchmendy (1)
TRW System Group
Space Park Drive
Houston, Texas 77058

Mr. Cecil Messer (1)
S&D - Dir
Marshall Space Flight Center
Huntsville, Alabama 35812

Mr. Fred Billingsley (1)
Code 168-421
Jet Propulsion Laboratory
4800 Oak Drive
Pasadena, California 91103

Ms. Retha Shirkey/JM6 (4)
NASA/JSC
Technical Library Branch
Houston, Texas 77058

Mr. W. Stoney/ ER (1)
Mr. Leonard Jaffe/ER (1)
Mr. M. Molloy/ERR (1)

NASA Headquarters
Washington, D.C. 20546

Dr. D. M. Deerwester
Ames Research Center
National Aeronautics and Space Administration
Moffett Field, California 94035

31F